CS 134
Problem Set 3

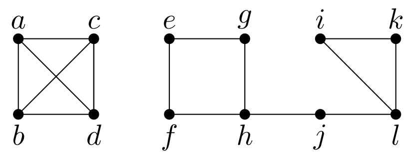
Xiner Zhou

February 28, 2017

1. Centrality (Jackson 2.2.4).

a. Compute the degree centrality of each of the nodes in network B.

Solution:



The nodes a, b, c, d, h, l all have degree 3, and the nodes e, f, g, i, j, k all have degree 2. Therefore,

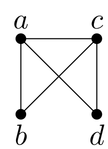
$$degree\ centrality(a) = degree\ centrality(b)$$
$$= degree\ centrality(c) = degree\ centrality(d)$$
$$= degree\ centrality(h) = degree\ centrality(l)$$
$$= \frac{3}{12-1} = \frac{3}{11}$$

$$degree\ centrality(e) = degree\ centrality(f) = degree\ centrality(g)$$
$$= degree\ centrality(i) = degree\ centrality(j) = degree\ centrality(k)$$
$$= \frac{2}{12-1} = \frac{2}{11}$$

b. Compute the closeness centrality of node d in network

$$(\{a, b, c, d\}, \{ab, ac, ad, bc, cd\}).$$

Solution:

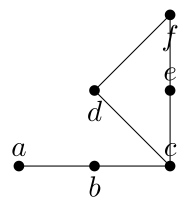


$$closeness\ centrality(d) = \frac{1}{\frac{dist(d,a)+dist(d,b)+dist(d,c)}{3}}$$

$$= \frac{1}{\frac{1+2+1}{3}}$$

$$= \frac{3}{4}$$

2. Betweenness Centrality. Compute the betweenness centrality of each of the nodes in the network $(\{a, b, c, d, e, f\}, \{ab, bc, cd, ce, df, ef\})$.



Solution:

- $betweenness\ centrality(a) = \frac{1}{10} \times$

$$(\frac{P_a(b,c)}{P(b,c)} + \frac{P_a(b,d)}{P(b,d)} + \frac{P_a(b,e)}{P(b,e)} + \frac{P_a(b,f)}{P(b,f)} + \frac{P_a(c,d)}{P(c,d)} + \frac{P_a(c,e)}{P(c,e)} + \frac{P_a(c,f)}{P(c,f)} \frac{P_a(d,e)}{P(d,e)} + \frac{P_a(d,f)}{P(d,f)} + \frac{P_a(e,f)}{P(e,f)})$$

$$= \frac{1}{10} \times (0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0) = 0$$

- $betweenness\ centrality(b) = \frac{1}{10} \times$

$$(\frac{P_b(a,c)}{P(a,c)} + \frac{P_b(a,d)}{P(a,d)} + \frac{P_b(a,e)}{P(a,e)} + \frac{P_b(a,f)}{P(a,f)} + \frac{P_b(c,d)}{P(c,d)} + \frac{P_b(c,e)}{P(c,e)} + \frac{P_b(c,f)}{P(c,f)} \frac{P_b(d,e)}{P(d,e)} + \frac{P_b(d,f)}{P(d,f)} + \frac{P_b(e,f)}{P(e,f)})$$

$$= \frac{1}{10} \times (1 + 1 + 1 + \frac{1}{2} + 0 + 0 + 0 + 0 + 0 + 0) = \frac{7}{20}$$

- $betweenness\ centrality(c) = \frac{1}{10} \times$

$$(\frac{P_c(a,b)}{P(a,b)} + \frac{P_c(a,d)}{P(a,d)} + \frac{P_c(a,e)}{P(a,e)} + \frac{P_c(a,f)}{P(a,f)} + \frac{P_c(b,d)}{P(b,d)} + \frac{P_c(b,e)}{P(b,e)} + \frac{P_c(b,f)}{P(b,f)} \frac{P_c(d,e)}{P(d,e)} + \frac{P_c(d,f)}{P(d,f)} + \frac{P_c(e,f)}{P(e,f)})$$

$$= \frac{1}{10} \times (0 + 1 + 1 + 1 + 1 + 1 + 1 + \frac{1}{2} + 0 + 0) = \frac{13}{20}$$

- $betweenness\ centrality(d) = \frac{1}{10} \times$

$$(\frac{P_d(a,b)}{P(a,b)} + \frac{P_d(a,c)}{P(a,c)} + \frac{P_d(a,e)}{P(a,e)} + \frac{P_d(a,f)}{P(a,f)} + \frac{P_d(b,c)}{P(b,c)} + \frac{P_d(b,e)}{P(b,e)} + \frac{P_d(b,f)}{P(b,f)} \frac{P_d(c,e)}{P(c,e)} + \frac{P_d(c,f)}{P(c,f)} + \frac{P_d(e,f)}{P(e,f)})$$

$$= \frac{1}{10} \times (0 + 0 + 0 + \frac{1}{2} + 0 + 0 + \frac{1}{2} + 0 + \frac{1}{2} + 0) = \frac{3}{20}$$

- $betweenness\ centrality(e) = \frac{1}{10} \times$

$$(\frac{P_e(a,b)}{P(a,b)} + \frac{P_e(a,c)}{P(a,c)} + \frac{P_e(a,d)}{P(a,d)} + \frac{P_e(a,f)}{P(a,f)} + \frac{P_e(b,c)}{P(b,c)} + \frac{P_e(b,d)}{P(b,d)} + \frac{P_e(b,f)}{P(b,f)} \frac{P_e(c,d)}{P(c,d)} + \frac{P_e(c,f)}{P(c,f)} + \frac{P_e(e,f)}{P(e,f)})$$

$$= \frac{1}{10} \times (0 + 0 + 0 + \frac{1}{2} + 0 + 0 + \frac{1}{2} + 0 + \frac{1}{2} + 0) = \frac{3}{20}$$

- $betweenness\ centrality(f) = \frac{1}{10} \times$

$$(\frac{P_f(a,b)}{P(a,b)} + \frac{P_f(a,c)}{P(a,c)} + \frac{P_f(a,d)}{P(a,d)} + \frac{P_f(a,e)}{P(a,e)} + \frac{P_f(b,c)}{P(b,c)} + \frac{P_f(b,d)}{P(b,d)} + \frac{P_f(b,e)}{P(b,e)} + \frac{P_f(c,d)}{P(c,d)} + \frac{P_f(c,e)}{P(c,e)} + \frac{P_f(d,e)}{P(d,e)})$$

$$= \frac{1}{10} \times (0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + \frac{1}{2}) = \frac{1}{20}$$

3. Programming: Verifying the Friendship Paradox. For each social network, compute the value of

$$\frac{\frac{1}{|G|} \sum_{n \in G} (deg(n))}{\frac{1}{|G|} \sum_{n \in G} (\frac{1}{|N(n)|} \sum_{m \in N(n)} deg(m))}$$

with the average being taken across all nodes n in the network G, and where $N(n)$ is the neighborhood of node n.

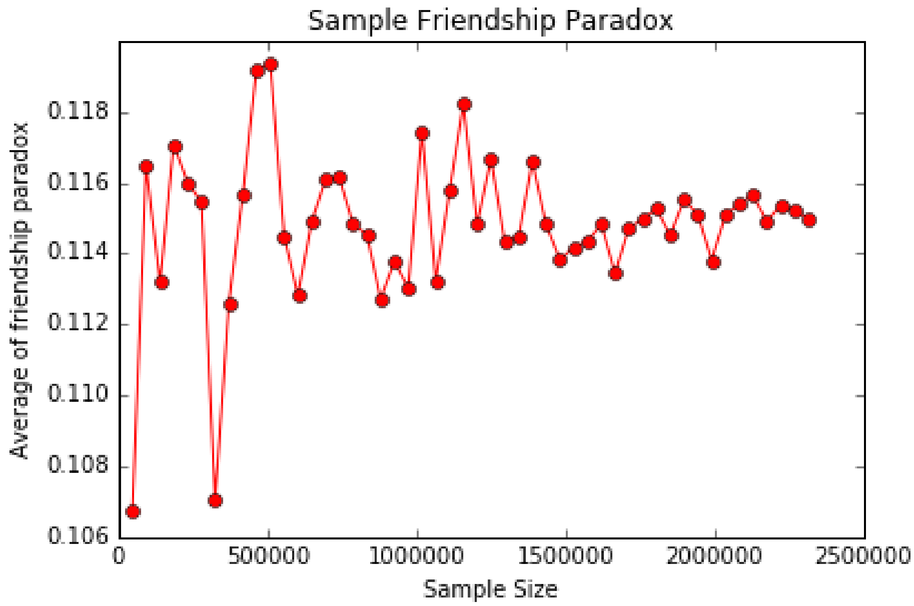
Solution: Since the networks are too large, I used the first 10,000,000 edges from each file, as a sample, to do the following analyses.

- Orkut social network and ground-truth communities (undirect): 0.002649
- LiveJournal social network and ground-truth communities (undirect): 0.041970
- Slashdot social network, November 2008 (directed): 0.114974
- DBLP collaboration network and ground-truth communities (undirected): 0.357449
- Enron email network (undirected): 0.042382
- Youtube social network and ground-truth communities (undirected): 0.008049
- Epinions social network (directed): 0.118562
- Wikipedia Talk network (directed): 0.001501

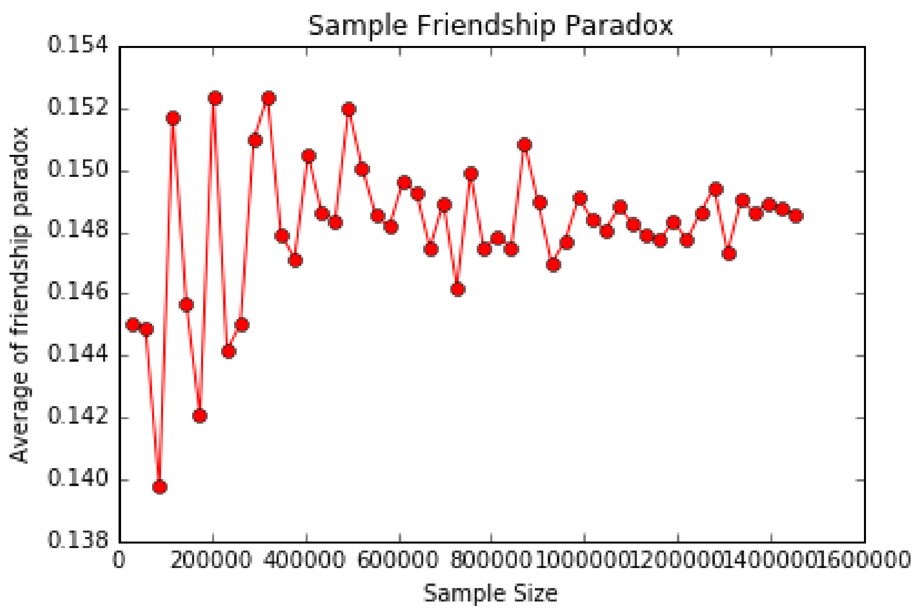
As a **bonus question (just for fun, no extra credit)**, for each network, plot the average of $\frac{deg(n)}{\frac{1}{|N(n)|} \sum_{m \in N(n)} deg(m)}$ against sample size (number of nodes n in the subsample). That is, for samples of varying sizes, compute this average for each sample on the y axis, and plot the sample size on the x axis.

Solution:

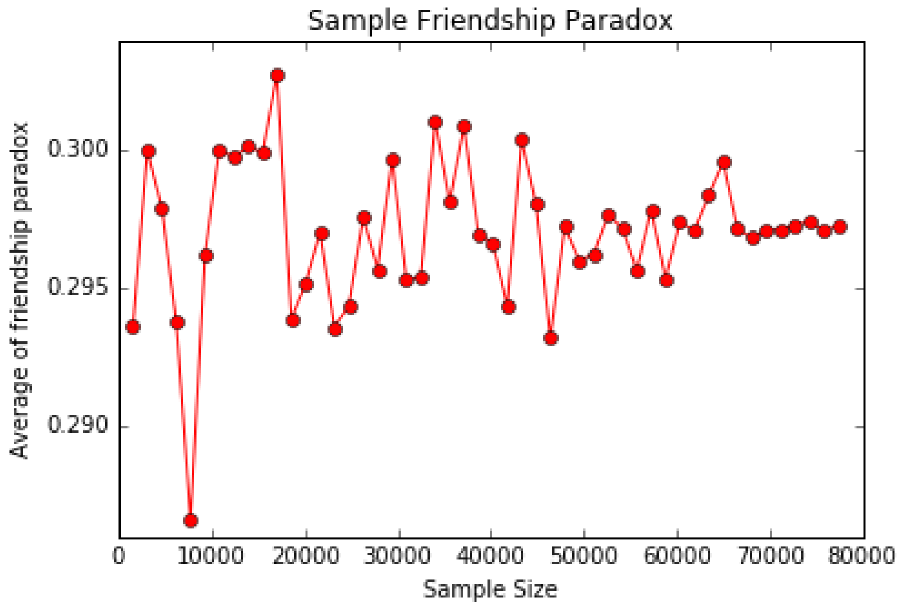
- Orkut social network and ground-truth communities (undirect):



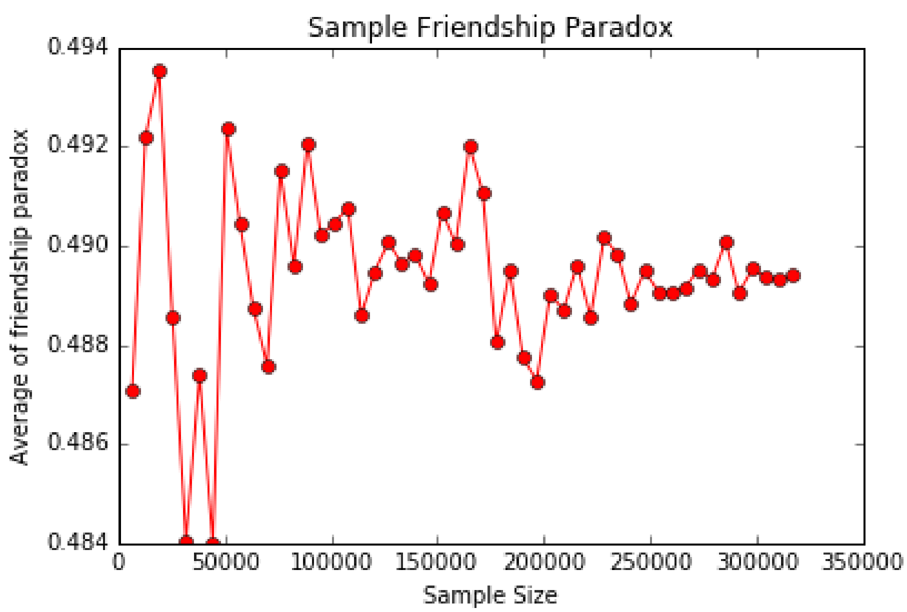
- LiveJournal social network and ground-truth communities (undirect):



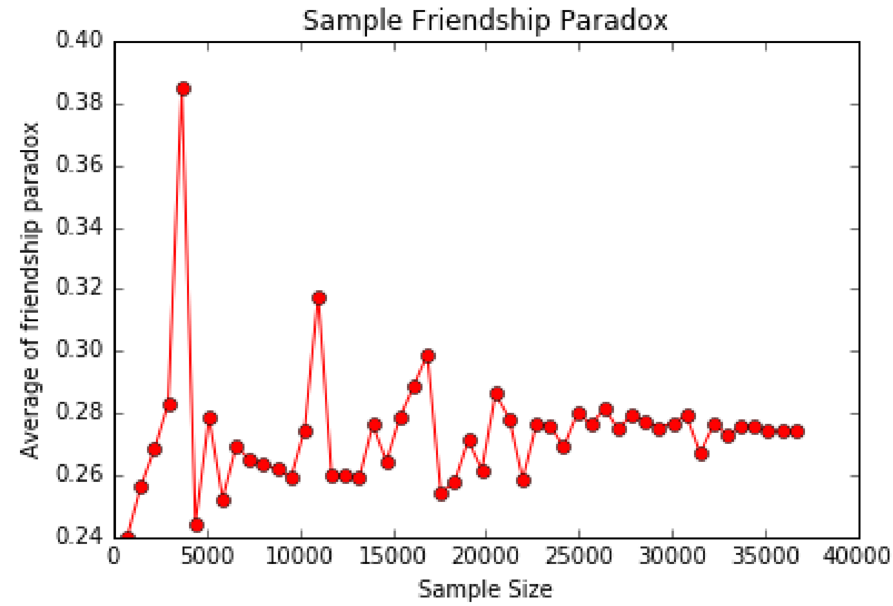
- Slashdot social network, November 2008 (directed):



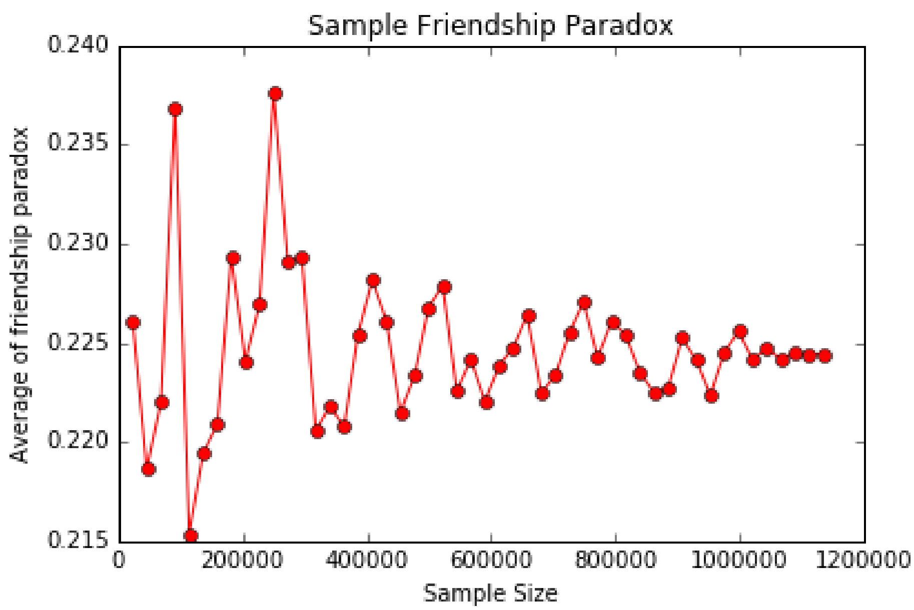
- DBLP collaboration network and ground-truth communities (undirected):



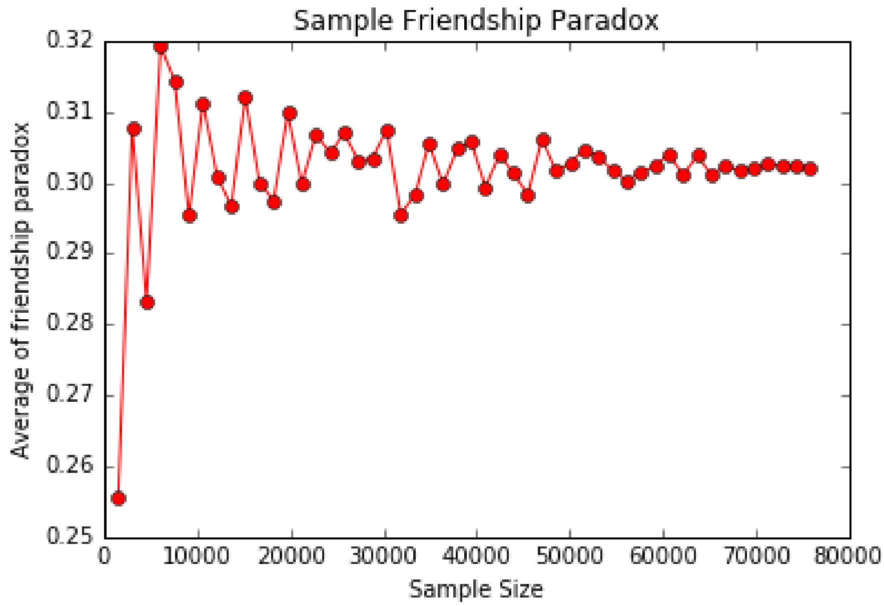
- Enron email network (undirected):



- Youtube social network and ground-truth communities (undirected):



- Epinions social network (directed):



- Wikipedia Talk network (directed):